DPPDRAGO.D71101

Inventor(s): Jean-Baptiste Dumas Milne Edy Aymeric Duclert, Lydie Bougueleret Our Ref.: 36.US3.DIV Express Mail Label No. EL821903714US

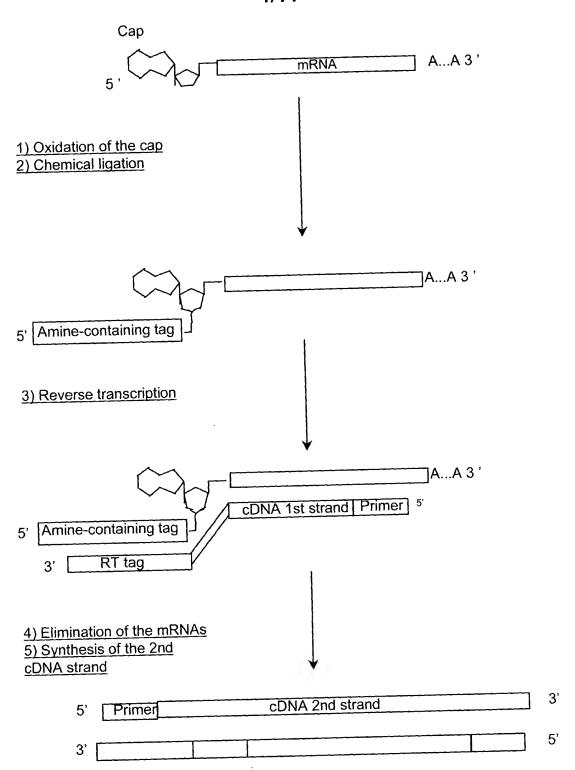


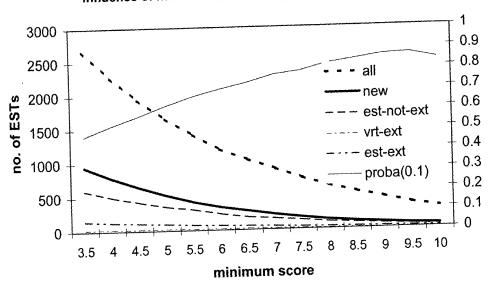
FIGURE 1

Aymeric Duclert, Lydie Bougueleret
Our Ref.: 36.US3.DIV
Express Mail Label No. EL821903714US
2/14

	Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
	3.5	0.121	0.036	0.467	0.004
- 1	4	0.096	0.06	0.519	0.664
	4.5	0.078	0.079	0.565	0.708
-	5	0.062	0.098	1	0.745
-	5.5	0.05	0.127	0.615	0.782
	6	0.04	0.163	0.659	0.813
	6.5	0.033	0.202	0.694	0.836
	7	0.025		0.725	0.855
	7.5	0.021	0.248	0.763	0.878
1	8	1	0.304	0.78	0.889
	8.5	0.015	0.368	0.816	0.909
	9	0.012	0.418	0.836	0.92
	9.5	0.009	0.512	0.856	0.93
		0.007	0.581	0.863	0.934
<u> </u>	10	0.006	0.679	0.835	0.919

3/14

# Influence of minimum score on signal peptide recognition



Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3.5	2674	947	599	23	150
4	2278	784	499	23	126
4.5	1943	647	425	22	112
5	1657	523	353	21	96
5.5	1417	419	307	19	80
6	1190	340	238	18	68
6.5	1035	280	186	18	60
7	893	219	161	15	48
7.5	753	173	132	12	36
8	636	133	101	11	29
8.5	543	104	83	8	26
9	456	81	63	6	24
9.5	364	57	48	6	18
10	303	47	35	6	15

FUI: COMPLEMENTARY DINAS

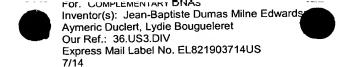
Inventor(s): Jean-Baptiste Dumas Milne Edwards
Aymeric Duclert, Lydie Bougueleret
Our Ref.: 36.US3.DIV
Express Mail Label No. EL821903714US
5/14

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	0
Dystrophic muscle	41	18	8	0	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	0
Large intestine	21	8	4	0	1
Liver	23	9	6	0	0
Lung	24	12	4	0	1
Lung (cells)	57	38	. 6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	0
Prostate	34	16	4	0	2
Spleen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	0
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

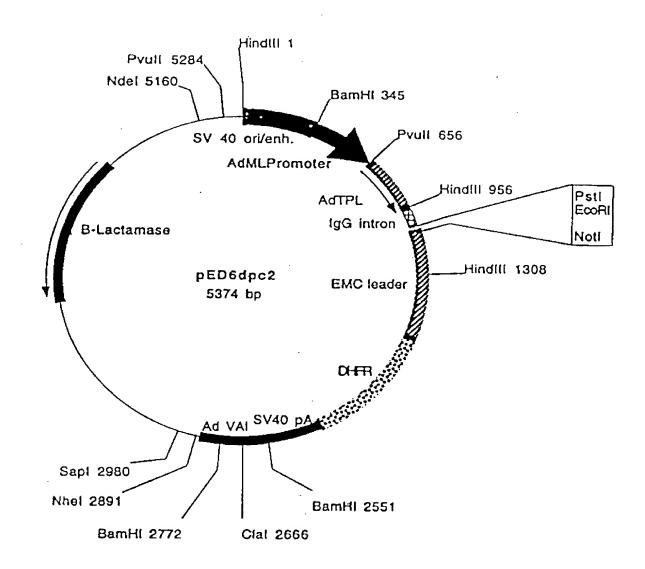
Express Mail Label No. EL821903714US 6/14 0-AAAReverse AAATTT transcription Nested PCR If Design of If 5' nested primers a) b) Primer walking and ORF Direct isolation Design of new Cloning primers TTT**PCR** 5 'EST Cloning ORF Signal peptide PCR primers AAA mRNA TTT. 1st strand. cDNA TTT PCR product Clone FIGURE 6

For. Complementary DNAS Inventor(s): Jean-Baptiste Dumas Milne Edw Aymeric Duclert, Lydie Bougueleret

Our Ref.: 36.US3.DIV



#### 7/14



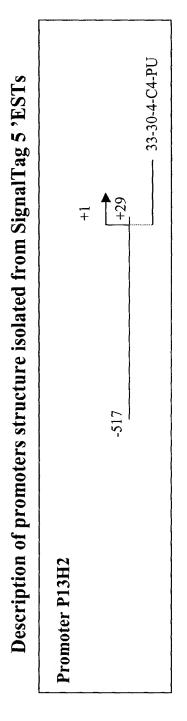
Plasmid name: pED6dpc2

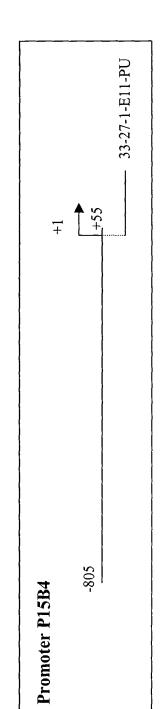
Plasmid size: 5347 bp

Comments/References: pED6dpc2 is derived from pED6dpc1 by insertion of a new polylinker to facilitate cDNA cloning. SST cDNAs are cloned between EcoRI and Not1. pED vectors are described in Kaufman et al. (1991), NAR 19:4485-4490.

For: COMPLEMENTARY DNAs
Inventor(s): Jean-Baptiste Dumas Milne Edward
Aymeric Duclert, Lydie Bougueleret
Our Ref.: 36.US3.DIV
Express Mail Label No. EL821903714US
8/14

8/14





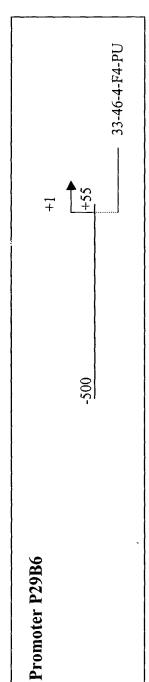


FIGURE 8

For: COMPLEMENTARY DNAs
Inventor(s): Jean-Baptiste Dumas Milne Edwar
Aymeric Duclert, Lydie Bougueleret
Our Ref.: 36.US3.DIV
Express Mail Label No. EL821903714US
9/14

9/14

#### Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences

Promoter sequence P13H2 (546 bp):								
Matrix	Position	Orientation	Score	Length	Sequence			
CMYB 01	-502	+	0.983	9	TGTCAGTTG			
MYOD Q6	-501	-	0.961	10	CCCAACTGAC			
S8 01	-444	-	0.960	11	AATAGAATTAG			
S8 01	-425	+	0.966	11	AACTAAATTAG			
DELTAEF1 01	-390	-	0.960	11	GCACACCTCAG			
GATA C	-364	-	0.964	11	AGATAAATCCA			
CMYB 01	-349	+	0.958	9	CTTCAGTTG			
GATA1 02	-343	+	0.959	14	TTGTAGATAGGACA			
GATA C	-339	+	0.953	11	AGATAGGACAT			
TAL1ALPHAE47 01	-235	+	0.973	16	CATAACAGATGGTAAG			
TAL1BETAE47 01	-235	+	0.983	16	CATAACAGATGGTAAG			
TAL1BETAITF2 01	-235	+	0.978	16	CATAACAGATGGTAAG			
MYOD Q6	-232	-	0.954	10	ACCATCTGTT			
GATA1 04	-217	_	0.953	13	TCAAGATAAAGTA			
IK1 01	-126		0.963	13	AGTTGGGAATTCC			
IK2 01	-126		0.985	12	AGTTGGGAATTC			
CREL 01	-123		0.962	10	TGGGAATTCC			
GATA1 02	-96	· · · · · · · · · · · · · · · · · · ·	0.950	14	TCAGTGATATGGCA			
SRY 02	-41	-	0.951	12	TAAAACAAAACA			
E2F 02	-33	+	0.957	8	TTTAGCGC			
MZF1 01	-5	-	0.975	8	TGAGGGGA			

For: COMPLEMENTARY DINAS Inventor(s): Jean-Baptiste Dumas Milne Edward Aymeric Duclert, Lydie Bougueleret Our Ref.: 36.US3.DIV Express Mail Label No. EL821903714US 10/14

10/14

Promoter sequence P15B4 (861bp) :								
Matrix	Position	Orientation	Score	Length	Sequence			
NFY_Q6	-748	-	0.956	11	GGACCAATCAT			
MZF1_01	-738	+	0.962	8	CCTGGGGA			
CMYB_01	-684	+	0.994	9	TGACCGTTG			
VMYB_02	-682	-	0.985	9	TCCAACGGT			
STAT_01	-673	+	0.968	9	TTCCTGGAA			
STAT_01	-673	_	0.951	9	TTCCAGGAA			
MZF1_01	-556	-	0.956	8	TTGGGGGA			
IK2_01	-451	+	0.965	12	GAATGGGATTTC			
MZF1_01	-424	+	0.986	8	AGAGGGGA			
SRY_02	-398	_	0.955	12	GAAAACAAAACA			
MZF1_01	-216	+	0.960	8	GAAGGGGA			
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC			
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC			
S8_01	5	_	0.992	11	GAGGCAATTAT			
MZF1_01	16	-	0.986	8	AGAGGGGA			

FIGURE 9 (cont)

FOI. COMPLEMENTARY DINAS Inventor(s): Jean-Baptiste Dumas Milne Edwa Aymeric Duclert, Lydie Bougueleret Our Ref.: 36.US3.DIV Express Mail Label No. EL821903714US 11/14

## 11/14

Promoter sequence P29B6 (555 bp) :								
Matrix	Position	Orientation	Score	Length	Sequence			
ARNT_01	-311	+	0.964	16	GGACTCACGTGCTGCT			
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG			
USF_01	-309	+	0.985	12	ACTCACGTGCTG			
USF_01	-309	-	0.985	12	CAGCACGTGAGT			
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT			
MYCMAX_02	-309	_	0.972	12	CAGCACGTGAGT			
USF_C	-307	+	0.997	8	TCACGTGC			
USF_C	-307	-	0.991	8	GCACGTGA			
MZF1_01	-292	-	0.968	8	CATGGGGA			
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT			
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC			
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC			
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC			
PADS_C	45	+	1.000	9	твтввтстс			

FIGURE 9 (cont)

Inventor(s): Jean-Baptiste Dumas Milne Edward Aymeric Duclert, Lydie Bougueleret Our Ref.: 36.US3.DIV Express Mail Label No. EL821903714US 12/14

12/14

97.8% identity in 92 aa overlap

			10	20	30	40	50	60
SEQ	ID	NO:120	MASLGHILVFCVG	<i>LLTMAKA</i> ESPKI	EH <b>DPFTYDYQ</b>	SLQIGGLVIA	GILFILGILI	VLSRRC
			::.: :::::::	:::::::::	::::::::::	:::::::::	:::::::::	:::::
SEQ	ID	NO:180	MAPLHHILVFCVG	LLTMAKAESPK	EH <b>DPFTYDYÇ</b>	SLQIGGLVIA	GILFILGILI	VLSRRC
			10	20	30	40	50	60
			70	80	90			
SEQ	ID	NO:120	RCKFNQQQRTGEP	DEEEGTFRSSI	RRLSTRRR			
			:::::::::::::::::::::::::::::::::::::::	:::::::::	:::::::			
SEQ	ID	NO:180	RCKFNQQQRTGEP:	DEEEGTFRSSI	RRLSTRRR			
			70	80	90			

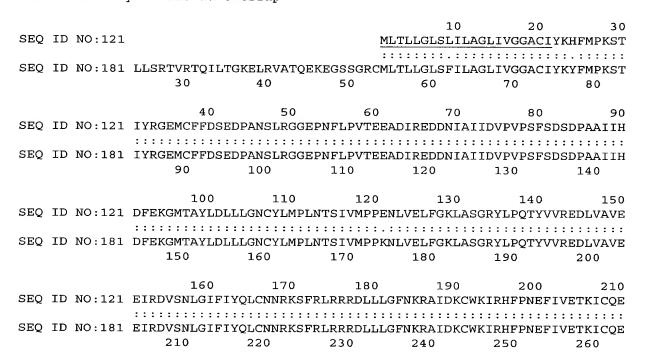
TOIL COMPLEMENTARY DINAS Inventor(s): Jean-Baptiste Dumas Milne Edward Aymeric Duclert, Lydie Bougueleret

Our Ref.: 36.US3.DIV

Express Mail Label No. EL821903714US

#### 13/14

98.6% identity in 210 aa overlap



SEQ ID NO:182 GTSHK



## 14/14

83.4% identit	y in 211 aa ov	erlap				
				10	20	30
SEQ ID NO:128	•		LWW	WLLWTVLILF	SCCCAFRHRR	AKLRLQ
			::::	:::::::::::::::::::::::::::::::::::::::	::::::::	:::::
SEQ ID NO:182	ELCPGVNTQPYLC	ETGHCCGETGC	CTYYYELWW	FWLLWTVLILF	'SCCCAFRHRR	AKLRLQ
	70	80	90	100	110	120
	40	50	60	70	80	90
SEQ ID NO:128	QQQRQREINLLAY	HGACHGAGPFF	PTGSLLDLRLI	LSTFKPPAYED	VVHRPGT <b>PPP</b>	PYTVAP
	:::::::::::::::::::::::::::::::::::::::				:::::::::	:::::
SEQ ID NO:182	QQQRQREINLLAY	HGACHGAGPVF	TGSLLDLRLI	SAFKPPAYED	VVHHPGTPPP	PYTVGP
	130	140	150	160	170	180
	100	110	120	130	140	150
SEO ID NO:128	GRPLTASSEQTCCS	SSSSSCPAHFE	EGTNVEGVSSH	IOSAPPHOEGE	PGAGVTPAST	PPSCRY
~						
SEO ID NO:182	GYPWTTSSECTRCS	SSESSCSAHLE	EGTNVEGVSSO	OSALPHOEGE	PRAGLSPVHI	PPSCRY
	190	200	210	220	230	240
		200				
	160	170	180	190	200	210
SEO ID NO:128	RRLTGDSGIELCP					
0-2 1- 10.110	::::::::::::::				- <del>-</del>	
SEO ID NO:182	RRLTGDSGIELCPO					
020 12 1101102	250	260	270	280	290	300
	250	200	270	200	200	300
SEO ID NO:128	CDID					
DEG TE MO.120	ODIE					